Las siete pautas para identificar y mitigar las campañas de ‘phishing’ basadas en IA

El phishing siempre ha sido una espina clavada en el costado de la ciberseguridad empresarial y los recientes desarrollos de inteligencia artificial como ChatGPT están empeorando aún más las cosas. Aquí dejamos algunas pautas para hacer frente a la cada vez más sofisticada amenaza del phishing.

La aparición de eficaces herramientas de procesamiento del lenguaje natural, como ChatGPT, significa que es hora de empezar a entender cómo protegerse contra los ciberataques basados en inteligencia artificial (IA). Las capacidades de generación de lenguaje natural de los grandes modelos de lenguaje (LLM) encajan a la perfección con uno de los vectores de ataque más importantes de la ciberdelincuencia: el phishing. El phishing se basa en engañar a la gente, y la capacidad de generar lenguaje efectivo y otros contenidos a escala es una herramienta importante en el kit del hacker.

Afortunadamente, existen varias formas de mitigar esta creciente amenaza. Aquí proponemos siete pautas para estar preparado en la era del phishing basado en IA:

Comprender la amenaza

Un líder encargado de la ciberseguridad puede adelantarse a los acontecimientos comprendiendo en qué punto nos encontramos en la historia del aprendizaje automático como herramienta de hackeo. En la actualidad, el área de mayor relevancia en torno a la IA para la ciberseguridad es la generación de contenidos. Aquí es donde el aprendizaje automático está haciendo sus mayores avances y encaja perfectamente para los hackers con vectores como el phishing y los chatbots maliciosos. La capacidad de elaborar textos convincentes y bien formados está en manos de cualquiera que tenga acceso a ChatGPT, y eso es básicamente cualquiera con una conexión a Internet.

“Buscar la mala gramática y la ortografía incorrecta es cosa del pasado; incluso los correos electrónicos de phishing anteriores a ChatGPT se han vuelto más sofisticados” afirmó especialista de coberseguridad.

“Debemos preguntarnos: ‘¿Es esperado el correo electrónico? ¿Es legítima la dirección del remitente? ¿Te incita a hacer clic en un enlace? La formación en concienciación de seguridad sigue teniendo un papel que desempeñar aquí”.

Una investigación de una empresa de ciberseguridad demuestra una serie de interacciones con ChatGPT en las que la IA genera correos electrónicos de phishing eficaces. Esta y otras investigaciones confirman lo que nosotros mismos podemos afirmar, que los raíles de seguridad destinados a impedir que las herramientas de IA se utilicen con fines ilegales no son fiables y que se están creando herramientas a medida para estos fines.

Debemos reconocer que la IA se puede utilizar ahora para generar contenidos eficaces y que va a mejorar en ello. Las herramientas LLM mejorarán, estarán más al alcance de los hackers y se crearán herramientas personalizadas para ellas. Ahora es un buen momento para empezar a pensar y tomar medidas para reforzar las políticas de seguridad.
También debemos esperar que el contenido de phishing no sólo sea más convincente, sino que esté mejor orientado, capaz de incorporar detalles específicos de tiempo, lugar y eventos. Los empleados ya no pueden confiar en los signos evidentes de que un correo electrónico es malicioso. Las imágenes, e incluso el audio y el video, pueden falsificarse con técnicas de generación de contenidos. Hay que reiterar continuamente que cualquier correo electrónico inesperado es sospechoso.

La mentalidad y la cultura son las principales defensas
"El 90% de las victimizaciones por ciberdelincuencia podrían evitarse fácilmente si los usuarios finales contaran con unos pocos conocimientos clave", explicó a CSO Scott Augenbaum, agente especial supervisor retirado de la División Cibernética del FBI. "¿Por qué no empezamos por ahí? Por desgracia, todo lo demás cuesta dinero y no parece funcionar. Ojalá alguien me dijera que estoy equivocado para poder jubilarme de verdad". "Tu primera línea de defensa es convertirte en tu propio firewall humano", dijo Augenbaum. Es decir, la mentalidad humana es la pieza central de la ciberseguridad. Por lo tanto, el cultivo de esa mentalidad dentro de una empresa es clave.

"La cultura se come a la estrategia para desayunar y siempre debe venir desde la alta dirección", afirmó Stu Sjouwerman, CEO de KnowB4. La mentalidad y el comportamiento cotidianos de los empleados constituyen el sistema inmunitario de base de la empresa, por lo que es fundamental formar a los empleados de forma sistemática para que sean conscientes de la seguridad". Con el phishing basado en IA, el mensaje importante es que no se debe dar importancia al correo electrónico ni a otras comunicaciones en función de lo pulido y sofisticado de su lenguaje. Los phishers ya no pasan la prueba de la risa y ahora se exige a los empleados un mayor grado de vigilancia.

Hacer hincapié en actuar correctamente
El correo electrónico y otros elementos de la infraestructura de software ofrecen una seguridad fundamental incorporada que garantiza en gran medida que no corremos peligro hasta que nosotros mismos tomamos medidas. Debemos ser muy conscientes de qué es lo que estamos haciendo. La información sensible no corre peligro hasta que un empleado envía una respuesta, ejecuta un archivo adjunto o rellena un formulario. El primer anillo de defensa en nuestra mentalidad debería ser: "¿Es legítimo el contenido que estoy viendo, no solo basándome en sus aspectos internos, sino teniendo en cuenta todo el contexto?". El segundo anillo de

defensa en nuestra mentalidad debe ser entonces: "¡Espera! Aquí me están pidiendo que haga algo".

Cuando los usuarios dan un paso más después de recibir un intento de phishing, eso es una gran victoria para los malos actores: sólo con ese elemento en su lugar puede proceder un ataque. Los profesionales de la seguridad deben formarse a sí mismos, a los empleados y a cualquiera que les escuche para que oigan las señales de alarma cuando se les pida que introduzcan información o ejecuten una aplicación desconocida.

Por supuesto, cuando se hace algo como transferir dinero, el sentido de la precaución debe ser elevado. Con los deepfakes, ha habido incluso casos de empleados que han creído que sus superiores les habían enviado indicaciones legítimas para enviar dinero. Las comunicaciones de gran importancia deben verificarse en un segundo canal no falsificable.

Realizar simulaciones de phishing

La única forma de ver lo bien que lo está haciendo una empresa para combatir el phishing es realizar pruebas. Realizar campañas de phishing con contenido generado por IA es una parte importante de la lucha contra esta amenaza. Ejecutar una campaña eficaz es un tema en sí mismo, pero la raíz de una buena campaña comienza con el establecimiento de objetivos concretos: las métricas que se pueden medir deben utilizarse para guiar las pruebas. Un buen ejemplo es medir la frecuencia con la que se denuncian los correos electrónicos de phishing y, a continuación, mover la aguja de ese indicador.

Crear una campaña contra el phishing también ayudará a comprender lo útiles que pueden ser las herramientas de IA para generar contenidos eficaces. Esto ayudará a reforzar la necesidad de tomarse el problema en serio. "Aunque la IA es persistente, es posible que tu seguridad sea resistente si refuerzas con frecuencia las mejores prácticas de seguridad y las pones a prueba", explicó especialista de seguridad. "Si actualmente no estás involucrando a tus empleados en ataques simulados de ingeniería social, ese es un gran elemento para agregar en un plan para 2023 para mejorar tu postura de seguridad y traer resiliencia a tu programa de seguridad".

Incorporar herramientas que automaticen la detección de IA

OpenAI (la empresa detrás de ChatGPT) y otras han lanzado herramientas para detectar texto generado por IA. Estas herramientas seguirán mejorando junto con los generadores de PNL, y pueden integrarse y automatizarse para ayudar a detectar contenido malicioso. Muchos proveedores de herramientas de análisis de correo electrónico están empezando a aprovechar la IA para ayudar a afinar la forma en que entienden contextos como los metadatos y la ubicación a la hora de evaluar qué es contenido legítimo. Combatir el fuego con fuego -en este caso, utilizar la IA para combatir la IA- es una parte importante del futuro de la ciberseguridad.

La detección del phishing es una parte clave de una estrategia global de red e infraestructura, y resulta especialmente eficaz cuando la infraestructura de reconocimiento e infiltración asistida por IA se combinan con la detección y la prevención asistidas por IA.

La detección de IA es una frontera activa en la investigación del aprendizaje automático. Esta investigación se seguirá llevando a la empresa como herramienta para luchar

contra el phishing basado en IA, y debería ser un espacio que vigilar de cerca en los próximos meses.

Proporcionar un mecanismo sencillo para denunciar el phishing

Alertar a la seguridad sobre el phishing es esencial para hacer frente a los ataques basados en IA. Dado que las campañas de IA pueden producirse en masa de forma más eficiente, es importante reconocerlas en el momento en que se desarrollan. Esto permite informar a los empleados rápidamente y proporciona datos fundamentales para las herramientas antiphishing y los modelos de detección de IA.

El informe de phishing es una parte vital de cualquier infraestructura de seguridad robusta y un informe eficaz se vuelve especialmente importante en el contexto de las campañas de IA debido a la mejora de la capacidad de los atacantes para escalar ataques de estilo spear-phishing (ataques que incorporan datos específicos desde dentro de la organización) mediante la automatización, recopilación e incorporación de dicha información. Este es un buen aspecto en el que centrarse cuando se realizan pruebas de los sistemas de detección y notificación de phishing.

Incorporar autenticación resistente al phishing

La autenticación basada en contraseñas es intrínsecamente susceptible al phishing, con técnicas como Captcha especialmente vulnerables a la IA. Por otro lado, existen métodos de autenticación resistentes al phishing. Las passkeys son probablemente el modo de autenticación más resistente al phishing. Todavía se están desarrollando e implantando, pero cada vez son más comunes. Una vez adoptadas, son básicamente inviolables.

La autenticación multifactor (MFA) también ayuda, porque la simple exposición de un nombre de usuario, contraseña combinada en un sitio de phishing o interacción no es suficiente para que un hacker obtenga acceso a un recurso si se requiere un autenticador secundario. CISA ha publicado una descripción general de las AMF resistentes al phishing.

-Matthew Tyson, cio.com
Autora: Mireya Cortés
Fuente de información: cio.com.mx